SUPEX : A Schema-Guided Path Index for XML Data

Jing Wang
Institute of Computing Technology, CAS, China
Renmin University of China

Supervisor : Prof. Shan Wang, Prof. Xiaofeng Meng
Overview

- Constructed based on DTD
- Integrating path index with value indexes
- Two main structures:
  - Structural Graph
    the structure summary of XML data
  - Element Map
    fast entries to nodes in Structural Graph
Architecture

Structural Graph

Element Map

DocId, Order:Size, Level, Info

Path Index Structure

SUPEX Structure With Value Indexes
The Procedure of Construction

- **Simplifying DTD**
  - E1* -> E1
  - E1? -> E1
  - E1+ -> E1
  - E1 | E2 -> E1, E2
  - ..., E1, ..., E1 -> ..., E1, ...

- **Constructing structural graph**
  Generate structural graph from DTD graph

- **Loading data**
  During XML document loading

(a) DTD Graph

(b) Structural Graph
Query Processing with SUPEX

- Two basic ways:
  - Given a tag, all elements with this tag can be obtained by the lookup of EM.
  - Simple label paths from the root of document can be matched by traversal of SG starting from the root node.

- Basic structural relationships:
  - E1/E2
  - E1//E2

- Partial label path:
  - //E1/.../En

- Query evaluation with value indexes
Preliminary Experiment Results

Element hierarchy of XMark Document

<table>
<thead>
<tr>
<th>Scaling Factor</th>
<th>Document Size (MB)</th>
<th>Element number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1.12</td>
<td>17132</td>
</tr>
<tr>
<td>0.05</td>
<td>5.6</td>
<td>83533</td>
</tr>
<tr>
<td>0.1</td>
<td>11.3</td>
<td>167865</td>
</tr>
<tr>
<td>0.5</td>
<td>56.2</td>
<td>832911</td>
</tr>
<tr>
<td>1.0</td>
<td>113</td>
<td>1666315</td>
</tr>
</tbody>
</table>

open_auction//description

description/text
Conclusion

- Generated from XML schema information
- Concise path index
- Integrating path index with value indexes
- Supporting label path expressions
- Supporting the evaluation of value-based condition predicates